

Revista Alicantina de Estudios Ingleses 23 (2010): 173-185

Violence in the Brothers Grimm's Fairy Tales: A Corpus-Based Approach

María Alcantud Díaz
University of Valencia
Maria.alcantud@uv.es

ABSTRACT

The purpose of this article is to carry out a corpus-based study on the presence of violence in a selection of eight tales by the Grimm's Brothers by looking at the terms which can be said to relate to the semantic field of violence. More specifically, this study will analyse a selection of eight tales in which the frequency of the words *cut*, *dead* and *blood* will be studied in detail. These words have been chosen due to their possible connection to violence after carrying out a quantitative analysis of the frequency of the whole main corpus. My initial hypothesis is that the corpus-based study of those eight tales would support my intuition regarding the high percentage of violence in the Brothers Grimm collection as opposed to the content of violence in a wider variety of texts. The study initially involved the analysis of frequencies of the lexical units in the Brothers' Grimm corpus and, secondly, a comparison of the results obtained in the frequency test to two reference corpora: the British National Corpus and the Cobuild Concordancer. The comparison and its results seem to indicate that there was a higher than average percentage of the use of words related directly or indirectly to violence.

1. Introduction

Kinder und Hausmärchen, known in English as *Children's and Household Tales* or *the Grimm's collection*) is a classic among children's literature. As children, we have all been fascinated by characters such as *Snow-White*, *Cinderella* and so on, and we have all dreamt of becoming one of them. Tatar (2004: xv), one of the worldwide renowned authorities on folklore and fairy tales, states in her preface to *The annotated Brothers Grimm* that "the fairy

tales collected in Germany almost two centuries ago by the Brothers Grimm continue to have a powerful hold on our culture. Adapted, revised, rescripted and bowdlerized, they greet us at the movies such as *Pretty Woman* or *Working Girl*, at the opera with *Hansel and Gretel* or *La Cenerentola*, and in advertisements for everything from Chanel to chocolate, and in visual media as often as in print”.

The purpose of this article is to carry out a corpus-based study on the presence of violence in a selection of eight tales by the Grimm’s Brothers, which compose the main corpus of the present study- by looking at the terms which can be said to relate to the semantic field of violence. More specifically, this study will analyse the frequency of the words *cut*, *dead* and *blood*. These words have been chosen due to their possible connection to violence after carrying out a pilot survey of the frequency of occurrence of the whole main corpus. For the present analysis, I understand violence as a verbal or non verbal action that results in physical or psychological harm to a person, object, place or animal. My initial hypothesis was that the corpus-based study of those eight tales would support my intuition regarding the high percentage of violence in the Brothers Grimm collection as opposed to the content of violence in a wider variety of texts. The study primarily involved the analysis of frequencies of the lexical units in the Brothers’ Grimm corpus and, secondly, a comparison of the results obtained in the frequency test with two reference corpora: British National Corpus and Cobuild Concordancer. The comparison and its results support the presence of higher- than-average percentage of the use of words related directly or indirectly to violence.

If my hypothesis was to be confirmed, the implications of my findings would be, that yet again, children are being exposed to violence through products which, supposedly, are products aimed at them.

Although this has been often affirmed and argued in numerous studies (i.e. Tatar 1987, 1992, 2004; Haase 2008; Zipes 1991;) My contribution may somehow be justified by bringing in a new view on the content of the brothers Grimm’s tales. That is, although many studies have paid attention to the violent content of the brothers Grimm’s tales, none, to my knowledge have supported their argument empirically. This is therefore the aim of this study, to fill an existing gap in the already numerous studies dedicated to the brothers Grimm’s tales. A corpus-based approach will provide a better understanding of such tales and make it possible for educators to deal with the tales more adequately. Labelling those readings more empirically and classifying in detail the lexical units belonging to the semantic field of violence can help understand the type of violence we are dealing with and, in addition, can help deal with it more effectively.

More specifically, this study will focus on the study of the frequency of occurrence of some of the lexical units referred to cruelty and violence found in eight out of the 210 tales which compose the brothers Grimm’s collection. These eight tales will be analyzed using Concapp, in order to prove that the content of violence in these tales is higher than average. The analysis will provide me with a frequency list of the lexical items which compose the Grimm’s corpus so as to highlight the percentages of some previously selected words related to cruelty and violence found in the tales. After that I will collate these results and compare them to other two corpora: the British National Corpus and the Cobuild Concordancer for comparison purpose. These results will be displayed in different graphics to make data understanding easier and visually clearer.

2. Outline of the present study

2.1 Statement of hypothesis

My main hypothesis here is that a corpus-based approach will contribute to a better understanding of such tales and will make it possible for educators to deal with the tales more adequately. Labelling the readings more empirically and classifying in detail the lexical units belonging to the semantic field of violence can help understand the type of violence we are dealing with and help deal with it more effectively.

2.2. Violence in the Brothers Grimm's Fairy Tales

As said before, the aim of the present study is to investigate the presence of violence in eight tales of the brothers Grimm's fairy tales collection using a corpus-based approach – in order to achieve an objective and empirical classification.

In my study, I primarily reviewed the different approaches which have been taken to research the presence of violence in the brothers Grimm's fairy tales, discovering that some of the most worldwide-recognised scholars who have researched the fairy tales, have provided many examples which confirm the high presence of violence in the Grimm's tales. Tatar, for instance confirms that “in fairy tales, nearly every character—from the most hardened criminal to the Virgin Mary— is capable of cruel behaviour” (1987:3-4), referring to the Grimm's tales. Interestingly, tales like *Hansel and Gretel*, *The boy who went forth to learn what fear was* and *The Juniper tree*, which all belong to the Grimm's collection, were included by Warner (1998:4), in a book described by the author herself as a book about fear. Not only violence, but the fact that the brothers Grimm tales were amongst the recommended titles by Nazis had a historical consequence: the Allied forces, after World War II, thought that the Grimm's fairy tales had contributed to Nazi atrocities and savagery. In fact, as Haase argues, the Nazis promoted German folk education and saw the folktales as a means to their racial and political ends (2008:407-408). For these reasons, their books were forbidden in England and America since, according to Haase, it was confirmed that these fairy tales were “profoundly repressive, fuelled prejudices and xenophobia, and glorified cruelty and militarism” (2008:408). Violence in the Grimm's tales is not only about killing or hurting, but also about indoctrinating male-children to learn fixed roles which have many factors in common with psychological or physical ill treatments of women by means of sexist and racist attitudes (Zipes 1991:47).

All the data mentioned above makes us question how it is possible to find so many scenes of violence in a collection of tales aimed at children. Tatar provides us with the answer to this question: the brothers Grimm had other scholars in mind when they published the tales (2004:xvi) as it was born as a philological research focused on collecting old stories from oral tradition—aimed at adults- to preserve German identity and not to achieve a collection of tales for children, in fact, the Grimm's collection became tales aimed at children due, at least in part, to marketing reasons (Alcantud 2009).

2.3 A corpus-based study of violence in the Brothers Grimm's Fairy Tales

Corpus linguistics focuses on the importance of studying patterns of real language research. More concretely, this discipline studies language through large collections of authentic –written and spoken- texts. These texts are called corpora and are used “to derive empirical knowledge about language” (Koteyko 2006:144-145). The above authors have highlighted the particular significance of *violence* in the brothers Grimm's collection from different approaches: socio-political, psychological and so on. Very little, however, has been investigated about the role played by violence within fairy tales from the point of view of a corpus-based approach. To fill this gap, I take as a starting point the work by Stubbs (1996), who explains how computer assisted analyses may provide a substantial and well documented alternative to the use of intuitive data as well as a new understanding of form-meaning relations. Stubbs establishes nine principles for this kind of analysis:

Linguistics is essentially a social science and an applied science; language should be studied in actual, attested, authentic instances of use, not as intuitive, invented, isolated sentences; the unit of study must be the whole texts; texts and texts types must be studied comparatively across text corpora; linguistics is concerned with the study of meaning: form and meaning are inseparable; there is no boundary between lexis and grammar: lexis and grammar are interdependent; much language use is routine; language in use transmits the culture; saussurian dualisms are misconceived (1996: 24-44).

The point of departure of my study will be Stubb's second and fourth principles, that is to say, that text must be studied comparatively across text corpora using authentic samples. As argued by Stubbs (1996: 45), when talking about Sinclairs' work, this

shows in a precise and concrete way how a large corpus and an associated technology create a viewpoint which can lead to innovations in linguistic description and theory. The essential vision underlying corpus linguistics is that computer-assisted analysis of language gives access to data which were previously unobservable, but which can now profoundly change our understanding of language”. Thus, using a computer assisted analysis in my analysis of the Grimm's Brother tales will enable me to prove my hypothesis in an innovative and objective way.

The present study will consider the frequency of use of some words related to cruelty and violence. Following Stubb's proposal (1996), I will take into account two factors in the study of the tales: the first one is the study of collocations of those words, extracted after doing the frequency test, which could be suspected of having some relation to cruel or violent situations. The second one will be the comparison of my findings to the same words in the reference corpora in order to study the percentages of use of those words in violent and cruel situations. As said before, a quantitative and empirical investigation of the Grimm's collection will be carried out using three distinct corpora:

Firstly, the 8 tales which compose a corpus of 17,416 words. An electronic version, collated to a printed version (Grimm, 2009), has been used to make corpora research by computer easier to use. These tales are:

9. *The Twelve Brothers*. Twelve brothers are obliged to leave the castle where they live and their little sister had to find them after suffering a fatal spell which made her remain mute.
11. *Little Brother and Little Sister*. Two children who were seriously ill treated by their step mother had to escape from their house. The boy suffers a terrible spell.
15. *Hansel and Grethel*. Two children, brother and sister, are abandoned by their father and kidnapped by a mean witch who tries to eat them.
16. *The Three Snake-Leaves*. A man who is willing to do anything to bring back his dead wife, and the hard price he must pay.
21. *Cinderella*. A girl who has no mother suffers with her father's new marriage to a widow with two other daughters. She is the object of serious ill treatments.
40. *The Robber Bridegroom*. A young bride is married to a murderer.
46. *Fitcher's Bird*. A wizard takes the role of a beggar and catches pretty girls.
53. *Little Snow-White*. A girl, whose mother died, sees how her step mother tries to kill her on several occasions simply due to envy.

Secondly, I have used the Cobuild Concordancer¹ corpus. In order to compare the percentages of frequency of some words related to violence which belonged to my main corpus (Grimm's tales), I typed in some simple queries (the words: cut, dead and blood) and I got a display of 50 concordance lines chosen at random from the corpus and which have been used for comparison purposes.

Thirdly I used British National Corpus.² I typed in the same queries as in Cobuild Concordancer, (cut, dead and blood) and the search result showed the total frequency in the corpus and up to 50 examples, used for comparison purposes as well.

I first carried out the quantitative analysis of the frequency by primarily identifying in a corpus all occurrences of a node word (word form or lemma) and its raw frequency. After that, I kept a record of collocates of this node which occurs in a window of defined size (i.e. four words to left and right). Then I counted the frequency of joint occurrence of node and each collocate; in this case the frequency of each collocation when related to violence and its estimation in percentages. Finally I collated these results with the ones achieved by the same process in the two reference corpora.

3. Methodology

In this section I describe the design of my study. I define the computational tool used to carry out my study and indicate the steps which have been followed. ConcApp Concordancing program³ is the program which has been selected for this task. As ConcApp web page argues, (it) "is a free and user-friendly text analysis program. It offers concordances, collocations and word frequency statistics. It can also be used to edit text files". ConcApp has been used in order to find the frequencies lists as well as the concordances lists in my tales corpus.

The corpus selected for the present study has already been presented as the brothers Grimm's tales. The method used for this analysis is based on computational corpus linguistics and comparative analysis. Comparative analysis implies carrying out an empirical analysis

of the main corpus, which involves a computational analysis of frequencies of the lexical units in the Brothers' Grimm corpus. For that reason I have generated a list of all the words in it, ordered in alphabetical order and by percentage of frequency, in which it is possible to observe the lexical units which are used in a higher percentage in the tales. Analyzing this kind of information provided me with: firstly a study of the most frequently used words- not taking into account function words. It enabled me to check the concordances of some words which, having a high percentage of frequency of use in the Grimm's corpus, are related to violent and cruel situations. In other words, it allowed me to find out when these words have been used in the Grimm's corpus relating to violence or cruelty.

Secondly, a comparative study of the same selected words. That is to say, after finding the percentages of frequency of use in violent or cruel situations in the Grimm's corpus, I collated them to the percentages of use in the same situations achieved when typing the same words in the other two reference corpora, the Cobuild Concordancer and the British National Corpus. These two corpora provided me with a random sample of 50 examples in which the queries asked were included. So it was possible to achieve the percentages of use of the selected words in violent situations. By doing this, and after collating all the results, it was possible to prove my hypothesis.

4. Results

My first step was to generate a list of the most frequently used words, not taking into account function words. By doing this I was able to check those which were directly or indirectly related to violent and cruel situations. I worked with just the first 200 words of the frequency list of the tales corpus. This list was cleared out as it is a well known fact that the words which are more frequently used in any kind of text are all those with a mainly grammatical meaning. Those words are pronouns, prepositions, articles and some others. I decided to exclude them from my list mainly since they exert no influence on the final results of my research (cf. Pérez Paredes 2002).

Thus after having removed all the function words, the list was composed of 89 words (see table 1)

WORD	FREQ	%	Queen	43	0,2465 %
Said	172	0,9862 %	Saw	43	0,2465 %
Little	113	0,6479 %	Come	41	0,2351 %
Came	76	0,4358 %	Sister	41	0,2351 %
Went	76	0,4358 %	Old	39	0,2236 %
Woman	52	0,2981 %	Great	37	0,2121 %
Have	51	0,2924 %	Cried	35	0,2007 %
King	48	0,2752 %	Maiden	34	0,1949 %
Go	46	0,2637 %	Looked	33	0,1892 %
Forest	44	0,2523 %	Door	32	0,1835 %
House	44	0,2523 %	Home	32	0,1835 %
Took	44	0,2523 %	Let	32	0,1835 %
			Beautiful	30	0,1720 %

Answered	29	0,1663 %	Pick	21	0,1204 %
Thought	29	0,1663 %	See	21	0,1204 %
Dear	28	0,1605 %	Daughter	20	0,1147 %
Time	28	0,1605 %	Brothers	19	0,1089 %
Day	27	0,1548 %	Brother	19	0,1089 %
White	27	0,1548 %	Began	19	0,1089 %
Children	26	0,1491 %	Cut	19	0,1089 %
Long	26	0,1491 %	Dead	19	0,1089 %
Put	26	0,1491 %	Evening	19	0,1089 %
Father	25	0,1433 %	Get	19	0,1089 %
Heard	25	0,1433 %	Opened	19	0,1089 %
Heart	25	0,1433 %	step-mother	19	0,1089 %
Son	25	0,1433 %	Sat	19	0,1089 %
Way	25	0,1433 %	Stood	19	0,1089 %
Child	24	0,1376 %	Wife	19	0,1089 %
Good	24	0,1376 %	Called	18	0,1032 %
king's	24	0,1376 %	Fell	18	0,1032 %
Art	23	0,1319 %	Girl	18	0,1032 %
Bride	23	0,1319 %	Tree	18	0,1032 %
Dwarfs	22	0,1261 %	Blood	17	0,0975 %
looking-glass	22	0,1261 %	Pieces	17	0,0975 %
Young	22	0,1261 %	Red	17	0,0975 %
Bread	21	0,1204 %	Ran	17	0,0975 %
Gave	21	0,1204 %	Fire	16	0,0917 %
Lay	21	0,1204 %	Find	16	0,0917 %
Like	21	0,1204 %	Got	16	0,0917 %
Night	21	0,1204 %	Has	16	0,0917 %
Poor	21	0,1204 %	Left	16	0,0917 %
			Mother	16	0,0917 %
			Morning	16	0,0917 %
			Man	16	0,0917 %
			Take	16	0,0917 %
			Asked	15	0,0860 %
			Ashes	15	0,0860 %
			Bird	15	0,0860 %
			Life	15	0,0860 %
			Pretty	15	0,0860 %
			Shoe	15	0,0860 %
			Through	15	0,0860 %

Table 1. Cleared frequency list.

Some other examples can be seen in lexical units related to the semantic field of ‘parts of the body’. The word *heart*, for instance, provided me with interesting data: almost every time this word was used, it was related to a violent or cruel situation (figure 3).

13 out of the 25 times this word is used in the text are related to violence. For example line

4 "If all this had been Richard he would not have been so
 5 poor, and so well off, and he would not have been so
 6 "poor," and so well off, and he would not have been so
 7 "poor," and so well off, and he would not have been so
 8 "poor," and so well off, and he would not have been so
 9 "poor," and so well off, and he would not have been so
 10 "poor," and so well off, and he would not have been so
 11 "poor," and so well off, and he would not have been so
 12 "poor," and so well off, and he would not have been so
 13 "poor," and so well off, and he would not have been so
 14 "poor," and so well off, and he would not have been so
 15 "poor," and so well off, and he would not have been so
 16 "poor," and so well off, and he would not have been so
 17 "poor," and so well off, and he would not have been so
 18 "poor," and so well off, and he would not have been so
 19 "poor," and so well off, and he would not have been so
 20 "poor," and so well off, and he would not have been so
 21 "poor," and so well off, and he would not have been so
 22 "poor," and so well off, and he would not have been so
 23 "poor," and so well off, and he would not have been so
 24 "poor," and so well off, and he would not have been so

Figure 3. Concordances of HEART.

6 “Came running by he stabbed it, and cut out its heart and took it to the Queen[...] or line 12
 “believing that she had eaten Snow-White’s heart, could not but thing she was the first[...]”.
 It has to be taken into account that this one is not a word related per se to it. These tables are
 a clear indicator of the content of the tales corpus, but my intention is to go one step forward
 in the study of the most frequent words.

My next step is to research the frequencies of three words selected from the list which, at first sight, are related to violence: *cut*, *blood*, *dead*. I have used ConcApp in order to find their frequency percentages. Figures 4, 5 and 6 show a detailed screenshot of ConcApp with the frequency percentages lists.

As it can be seen, there are 19 examples of the word *cut* which represents 0.1089% of the

Unique Words View

Unique words = 1937
Total Words = 1741

OK Cancel

Alphabetic Sort Frequency Sort

Unique Words	Instances	Percent
music	1	0.0057 %
matches	1	0.0057 %
my	2	0.0115 %
curiously	1	0.0057 %
curiosity	2	0.0115 %
curve	1	0.0057 %
curtains	1	0.0057 %
cut	19	0.1389 %
cutting	1	0.0057 %

Figure 4. Cut.

Unique Words View

Unique words = 1937

Total Words = 12441

OK Save

Alphabetical Sort Frequency Sort

Unique Words	Instances	Percent
daughter	20	0.1147 %
daughters	3	0.0172 %
daunted	1	0.0057 %
day	27	0.1548 %
daybreak	1	0.0057 %
days	4	0.0229 %
dead	19	0.1089 %
deadly	1	0.0057 %
Dear	28	0.1625 %

Figure 5. Dead.

Unique Words View

Unique words = 1382
Total Words = 23260

OK Save

Alphabetic Sort Frequency Sort

Unique Words	Instances	Percent
bleased	2	0.0086 %
sleazebag	2	0.0086 %
blindness	2	0.0086 %
sleaz	3	0.0129 %
blood	18	0.0774 %
sleazebait	2	0.0086 %
sleazy	4	0.0172 %
sleazebag	2	0.0086 %
sleazebag	2	0.0086 %

Figure 6. Blood.

whole corpus. In terms of the word *blood*, there are 18 examples (0.0774 %), which should be added to the two examples of the word *blood-red* (0.0086%) and the 4 examples of the word *bloody* (0.172%). Finally, I have found 19 examples of the word *dead* (0.1089%) plus one example of the word *deadly* (which is not representative), as illustrated in table 2 below.

Selected words	Frequency percentage	Figure
CUT	0.1089 %	7
BLOOD	0.0774 %	8
DEAD	0.1089 %	9

Table 2. Frequency percentages of cut, blood and dead.

These percentages can be considered as high percentages bearing in mind we are working with a list composed of the 89 most frequent words in the tales corpus. This table confirms once again the remarkable presence of violence in these tales, since they are words related intrinsically to this semantic field, but are all these examples used violently in the corpus? In order to answer this question, my next step will be to extract percentages of how many of times these words have been used related to violence and cruelty within the tales by using a concordance analysis. In figures 7, 8, 9, 10, and 11 screenshots of the concordances related to these words in my corpus will be seen:

Let's study the times in which these words appear in violent or cruel situations in table 4:

1 "But what did she see when she went in? A great bloody basin stood in the middle of the room, and therein
2 as once by the red spots that she had seen in the bloody chamber. "Since then, she had gone into the room again,
3 and away by her curiosity, opened the door of the bloody chamber, looked in, and had to stare at it with a
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144

	DEAD	CUT	BLOOD
Total numb. of examples	19	19	24
Numb. of times in violent situations	19	12	15
Percentages	100%	63.1%	62.5%

Table 3. Percentages of frequency

As we can see in table 3, a high percentage of times (100% for *dead*) these words have been used in violent situations. These are very high percentages and they might already confirm my hypothesis which stated the presence of a high percentage of violence in the Grimm's tales. However once I have studied the corpus, my next step is to look for the same words in the two corpora proposed for comparison purpose: BNC, The British National Corpus and Cobuild Concordancer. The reason for this comparison is to determine if the percentages of frequency of use of these words are higher or shorter than the ones found in the reference corpora.

The next step in the analysis involves generating examples of the use of these three words in a larger corpus; with the intention to compare them to the corpus of the brothers Grimm's tales.

First a random selection of 50 solutions was taken from the British National corpus and the Cobuild Concordancer. The results are illustrated in tables 4 and 5 below:

	DEAD	CUT	BLOOD
Samples	50	50	50
Times in violent situations	17	3	9
Percentages	34%	6%	18%

Table 4. British National Corpus.

	DEAD	CUT	BLOOD
Samples	50	50	50
Times in violent situations	14	0	6
Percentages	22%	0%	12%

Table 5. Cobuild Concordancer.

If we combine all the percentages in frequencies related to violent situations in just one table (table 6):

	Cut	Blood	Dead
Grimms Tales	63.1%	62.5%	100%
BNC	6%	18%	12%
Cobuild Concordancer	0%	12%	22%

Table 6. Comparison British National Corpus, Cobuild Concordancer and Grimm's corpora.

And we transform these data into a graph in order to obtain visual evidence of the difference in percentages (see graphic 1), we can observe that the Brother Grimm's corpus percentages on the use of violence are largely higher than the results obtained in the two reference corpora used for comparison purposes.

5. Conclusions

In this research into the brothers Grimm's fairy tales, I have studied a corpus composed of 8 of their tales in order to make a twofold computational analysis. This study has been carried out in two steps: firstly the analysis of frequencies of the lexical units in the Brothers' Grimm corpus. Secondly, comparing it to the British National Corpus (BNC) and the Cobuild Concordancer in which I found a high percentage of use of words related directly or indirectly to violence.

The computational study unquestionably provided confirmation of a high percentage of violence in the brothers Grimm's tales. Thus, I have found out the frequency of use and percentages of some words related to cruelty and violence: cut + parts of the body, dead and blood found in the tales. After that I have compared these results to other two corpora, the British National Corpus and the Cobuild Concordancer, in order to discover such a high percentage of violence in the text which leads us to think of them as not aimed at children. It is widely recognized, violence can be found where you least expect it: cartoons, sports, readings, movies, etc and the presence of violence in the Grimm's collection is another key example.

The true aim of this article was to demonstrate that classical tales aimed at children can be relabelled by way of a corpus-based approach in an objective and empirical way, taking into account the evolution of society. In this sense it seemed worthwhile using the presence of violence in the brothers Grimm's fairy tales collection to detect it following Stubb's line of study on corpus analysis. Thus, my point of departure was to demonstrate a higher than average percentage of presence of actions related to violence and cruelty within the Grimm's corpus empirically. After doing it, it has become apparent that any lexical unit which might be controversial – in other words, topics which could not be total or at least in part aimed at children- might be detected.

Almost no investigation has been carried out by means of a computer-based approach to the brothers Grimm's fairy tales. My results seem to be pointing to the possibility of getting classical readings relabelled according to a standard, empirical and objective list of conventions on literature aimed at children.

The present research is just a starting point of what I hope will offer a thorough insight into the Brothers Grimm's Fairy tales. I have only covered eight out of the 209 tales taken from the original version but, as observed, the presence of violent behaviour is a characteristic of almost all of them. The results, no doubt, are interesting enough to divert the public's attention towards a more in depth and multidisciplinary study of the whole collection itself, which I hope to continue doing. I realise the limitations in the review of the literature section which I hope to extend by including the perspectives of other disciplines such as psychology and social sciences, which, I am sure will give me a richer insight into the interpretation of the tales

themselves and how they are or should be used in nowadays society.

To finish with, I want to make it clear that the historical and literary quality of the brothers Grimm Collection is not being questioned in this article. The only point here is to study the possibility that this collection does not suit the scale of values that we are trying to instil in our children according to the society in which we are living nowadays.

Notes

1. <http://www.collins.co.uk/Corpus/CorpusSearch.aspx>.
2. <http://www.natcorp.ox.ac.uk/corpus/index.xml>
3. <http://www.edict.com.hk/PUB/concapp/> and
<http://www.intute.ac.uk/artsandhumanities/cgi-bin/fullrecord.pl?handle=humbul4918>

References

- Alcantud-Díaz, María (2009): *Cruelty and violence in the brothers Grimm's fairy tales collection: a corpus linguistics analysis*. Research project. Universitat de València.
- Atkins, B.T.S. and Zampolli, A.(1994): *Computational Approaches to the lexicon*. Oxford: Oxford University Press.
- Biber, D., Conrad, S. & Reppen, R. (1998): *Corpus linguistics investigating language structure and use*. Cambridge: Cambridge University Press.
- Grimm, Jacob.& Wilhelm. (2009): *The complete fairy tales, the Brothers Grimm*. Ware, Hertfordshire: Wordsworth Editions Limited.
- Grimm, Jacob & Wilhelm (trans. Margaret Hunt) (1884):*Household tales with the authors notes, translated by Margaret Hunt*.
[<http://www.surlalunefairytales.com/authors/grimms.html>].
- Haase, Donald. (2008): *The Greenwood encyclopedia of folktales & Fairy tales*. The USA: Greenwood Publishing Group.
- Koteyko, Nelya (2006): "Corpus linguistics and the study of meaning in discourse". *The linguistics Journal* 1(1): 132-157.
- Pérez Hernández, M. C. (2002): "Explotación de los corpórea textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento." *Estudios de Lingüística Española (ELiEs)* 18. [<http://elies.rediris.es/elies18/>].
- Pérez Paredes, Pascual (2003): "Small corpora as assisting tools in the teaching of English news language: A preliminary tokens-based examination of Michael Swan's Practical English Usage news language wordlist". *Journal ESP world*. [http://www.esp-rld.info/articles_6/issue_6.htm].
- Rinolucrí, Mario (2003): "The challenge corpus linguistics offers to language teachers". in G. Luque Agulló, A. Bueno González & G. Tejada Molina (eds), *Las lenguas en un mundo global – Languages in a global world*. Jaén: Universidad de Jaen, 147-156.
- Stubbs, M. (1996): *Text and corpus analysis*. Oxford: Blackwell.
- Tatar, Maria (1987):*The hard facts of the Grimms' Fairy tales*. Princeton: Princeton University Press.
- _____. (1992): *Off with your heads*. Princeton: Princeton University Press.
- _____. (2004): *The annotated Brothers Grimm*. New York: W.W. Norton & Company.
- Zipes, Jack (1991): *Fairy Tales and the Art of Subversion. The Classical Genre for Children and the Process of Civilization*. New York: Routledge.